Introduction to Network Meta-Analysis

CADTH Symposium 2013, NMA Workshop (Edward Mills, Kris Thorlund, Brian Hutton)
Brian Hutton, MSc, PhD
Associate Scientist & Senior Methodologist, Knowledge Synthesis Group, OHRI
Adjunct Professor, Dept. Of Epidemiology & Community Medicine, University of Ottawa
Outline for The Next Hour

• Brief Overview: Context for today

• RAPID reminder: Traditional systematic reviews/meta-analysis
  – What are they?
  – Why do we do them in healthcare research?

• Network meta-analysis (today’s workshop focus):
  – Conceptual overview (GAD example; Ioannidis; others)
  – Role in meta-analysis; advantages vs. standard MA; statistical set-up and available tools
  – ‘direct’ and ‘indirect’ evidence; what does this mean?
  – Brief assumptions overview (terminology):
    • homogeneity, similarity, consistency…what are these?
  – Reporting considerations for network meta-analysis
Context for today: NMA in Health Technology Assessment

- HTA, reimbursement decision-making (clinical & economics) commonly based on SR/MA of RCTs which compare competing trts.

- Increasingly common for diseases to have *multiple treatment options*.
  - May not have been compared in the context of RCTs.
    - Ethical issues, costs, time all reasons for lack of RCT data
  - Traditional pairwise meta-analysis methods don’t apply well
  - New drug ‘A’ needs to be compared to existing drug ‘B’…but only placebo-controlled trials exist. Cannot wait for a trial to be done!

Path forward?
Can trt A be compared to trt B despite an absence of ‘direct’ evidence?
• Sometimes more than two treatments will be considered. There may also exist both ‘direct’ and ‘indirect’ evidence.

• Both ‘simple’ indirect comparisons and network meta-analyses increasingly used/accepted by technology assessment agencies…
Effectiveness of smoking cessation therapies: a systematic review and meta-analysis

Ping Wu, Kumanan Wilson, Popey Dimoula and Edward J Mills

Address: 1Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada. 2Department of Medicine, University of Toronto, Toronto, Canada and 3Faculty of Health Sciences, Clinical Epidemiology & Biostatistics, HSC-2C12, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada
• Indirect comparisons and network meta-analyses are increasingly common (and important) in evidence synthesis for HTA.

• Both the ‘generators’ of evidence for HTA and reimbursement agencies need to develop greater understanding of this methodology.

• ***Awareness of key methods planning to maximize validity are critical***

• The objective of today’s session is to develop understanding for NMA, considerations for conduct, and insights on how to appraise them.
  • Highlight key areas of focus for ‘producers’
  • Highlight key areas of appraisal for ‘users’

• An enhanced understanding of valid methods by all (and how to appraise validity!) will lead to higher quality reviews and more efficiency in reimbursement decision-making.
‘Traditional’ Systematic Reviews: the Premise
You’re a clinician treating a patient with generalised anxiety disorder.

You are considering prescribing one of two pharmacotherapies: the SSRI paroxetine, or the SSRI sertraline.

You’re unsure which is more likely to balance effectiveness and safety for your patient. Is there evidence pertinent to making this decision???
What would be useful evidence here?

• “Hierarchy of evidence” tells us that if we could find a randomized controlled trial (RCT) comparing paroxetine with sertraline in the right type of patients, this would be our most unbiased evidence.

• Rather than just ONE RCT, what if we could find ALL OF THEM?
  • Could see if all studies show a similar “winner”
  • Could see if all studies show a similar “winning gap” (treatment effect)
  • Could calculate an ‘average’ treatment effect across the studies. More patients and studies = more precision for our estimate.

• The idea of finding ALL relevant RCTs and exploring these properties of existing experimental data is the premise of systematic reviews and meta-analysis.

• Such works are used by clinicians for evidence-based practice, and by a broad range of healthcare decision-makers to inform decisions.
Definitions:

**Systematic review (Cook, Sackett, Spitzer):**

- “A review in which there is a comprehensive search for relevant studies on a specific topic, and those identified are appraised and synthesized according to a predetermined method”

**Meta-analysis (Rothman, Greenland):**

- “A statistical analysis of a collection of studies”
- “Focus is contrasting and combining results from different studies to identify patterns, disagreements…”
- Estimate a ‘weighted average’ across relevant studies
- *Most prevalent use:* combine RCTs to get a pooled estimate of treatment efficacy/effectiveness relative to some comparator. Focus is on *two treatments.*
Idea: combine the 5 trials…

To get a summary estimate

Clinical conclusions about “A versus B” typically drawn on this summary estimate and consistency of findings from contributing studies
“Standard” PICO structure used for research questions in SR / MA:

<table>
<thead>
<tr>
<th>Does A produce better benefits than B in population X?</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g. some new drug therapy or other form of trt</td>
</tr>
<tr>
<td>e.g. mortality, physical function, reduced frequency of ???</td>
</tr>
<tr>
<td>e.g. current standard therapy, placebo, or other active comparator</td>
</tr>
</tbody>
</table>

However, how often does “A versus B” address question of interest?
- Many diagnoses with >2 therapies available.
- Stopping at 2 = “short-sighted view of the evidence”

***Need to modify approach to deal with A, B, C, D, E, etc to generate more clinically meaningful findings***. HOW?
Focus of Today’s Workshop: Network Meta-Analysis
Examples of Situations Needing ‘Modified’ Approach...

- Biologics for rheumatoid arthritis
  - Etanercept, infliximab, anakinra, abatacept, rituximab, certolizumab, golimumab, tocilizumab, adalimumab

- Treatments for smoking cessation
  - Pharmacotherapies (varenicline, bupropion, NRT gum, patch, inhaler)
  - Behavioral therapies, etc.

- Agents for prevention of secondary vascular events
  - ASA, dipyridamole, ticlopidine, clopidogrel, combinations thereof…

“A versus B” not always enough: need to go >2 therapies
If We Think Back to Generalized Anxiety Example...

• Just thinking about paroxetine and sertraline is too limited; there are many other pharmacotherapies available for GAD.
• An existing review with network meta-analysis assessed many other options (Baldwin BMJ 2011); makes for a much more clinically relevant question.

“Treatment Network Diagram” of treatments and available evidence
Another problem with standard “A versus B” approach...

- Placebo-controlled trials often suffice for regulatory approval; head-head RCTs of active interventions often non-existent.
  - ‘A versus B’ reviews may not address ‘right’ question, only comparisons with no therapy

- Decision-making cannot wait for the “perfect” evidence (which may never arrive)

- So then what? Must do the best with what’s available…
  - Proposed solution: employ placebo-controlled studies as data source to derive ‘indirect’ comparisons of active therapies (Ades et al)
  - Where available, supplement with head-head (‘direct’) data
  - i.e. combine ‘direct’ and ‘indirect’ evidence
**Terminology:**  
“Direct” and “Indirect” Evidence (compare drugs A, B)

Direct Comparison:

1 or more RCTs  
A → B

Estimates from RCTs, pairwise meta-analysis

Indirect Comparison:

1 or more RCTs  
A → C  
1 or more RCTs  
C → B  
0 RCTs

Can compare A, B via ‘common comparator’ C
A simple and silly (but helpful) example...  
Who’s the strongest celeb???

- Key candidates:
  - “All American Hero” Hulk Hogan
  - Sylvester “Rambo” Stallone
  - Arnie “the governator” Schwarzenegger

- All were on a celebrity arm wrestling circuit, so should be easy to check win/loss on the circuit’s CPU and find the answer, right?

- Problem: they never actually went head-to-head, only completed against one lesser, common opponent...
What we want to know

6W, 1L: WP=86%

11W, 3L: WP=79%

3W, 4L: WP=43%

What we know
How to choose strongest with no head-heads?

• Deduction?
  • E.g. compare winning percentages of Hogan, Stallone, Schwarzenegger versus Urkel to ‘choose a winner’?
  • Likely the approach if missing primary data

• Based on WP:
  • #1 = Hogan; #2 = Schwarzenegger; #3 = Stallone

• But some things to think about:
  • Was Urkel in the same physical shape for all matches? Rules always always the same? Etc.
  • What explains Stallone’s poor performance? Something other than strength?
Moving back to the realm of medical research...
What we know

What we want to know

5% v. 10% mortality

3.7% v. 8.5% mortality

7% v. 12% mortality
Consider indication X, for which there are 7 active therapies available, but only the following studies in published primary research.

**Placebo controlled ("indirect") data:**

- Drug A
- Drug B
- Drug C
- Drug D
- Drug E
- Drug F
- Drug G

**Head-to-head ("direct") data:**

- Drug A
- Drug B
- Drug C
- Drug D
- Drug E
- Drug F
- Drug G

Could be 6 or as many as 13 meta-analyses...

Ioannidis J. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009.
…or 1 network meta-analysis, which also fills in missing comparisons for us through indirect paths (think triangles)...

Ioannidis J. Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. CMAJ 2009.
Can work with any number of comparators and networks of different architectures if appropriate RCT data linking the treatments exists;

Past examples have been of varied complexity...
Biologics for Rheumatoid Arthritis (CADTH 2010)
Pharmacologic Agents for Depression (Cipriani et al 2009)

Treatments for Rheumatic Atrial Fibrillation (Sutton et al 2006)
Chemotherapy regimens for breast cancer (Mauri et al 2008)
Benefits of Using Network Meta-Analysis

• **Answer more policy relevant questions**
  - All therapies of interest versus a subset

• **Gain precision by considering all available evidence**
  - E.g. Might have only 1 study of A vs B, but 10 studies of A vs placebo and 20 of B versus placebo. More studies and patients can help narrow confidence intervals

• **Capable of helping to ‘pick a winner’ amongst all treatments**
  - Markov Chain Monte Carlo simulation method for estimation which allows us to draw thousands of iterations and estimate probabilities each treatment is the best
Patterns: Uptake in Use

Salanti (2011): rapidly expanding, with articles now appearing in highest impact journals including BMJ, JAMA and elsewhere. Now more than 200 published.

- Widely used in health technology appraisals around the world for various clinical indications with many trts (diabetes, obesity, rheumatoid arthritis, etc.)

- A new dedicated Cochrane methods group has formed
  - “Comparing multiple interventions methods group” (CMIMG)

- In Canada, CIHR recently awarded funds to groups of researchers to continue development of network MA methodology and applications.
  - Mills, Thorlund, Hutton part of the NETMAN funded team
Assumptions & Validity for Network Meta-Analysis: what do we need to consider?
The validity of NMA is discussed as relying on three assumptions:

1. **Homogeneity**
2. **Similarity**
3. **Coherence**
Trials on the same comparison: are the results homogeneous or heterogeneous?

Similarity

Coherence
Trials on the same comparison: are the results homogeneous or heterogeneous?

Trials across comparisons: are these trials similar enough to consider together?

Coherence
Trials on the same comparison: are the results homogeneous or heterogeneous?

Trials across comparisons: are these trials similar enough to consider together?

Network of many comparisons: are the results from direct and indirect comparisons consistent?
In practice, how to assess assumptions?

- **Homogeneity:** assessed within each comparison in the network; look at how much variation of treatment effects there is amongst the studies along each link (so $I^2$, for example, is helpful). Clinical homogeneity also reviewed...

Reasonable for us to combine studies of A vs B?

What about A vs placebo?

And How about B vs placebo?
Assessing Similarity in trials of RA MTX naïve pts

Across the treatment network, are trials similar to each other with the exception of the treatments being compared?

<table>
<thead>
<tr>
<th>Trial</th>
<th>Biologic</th>
<th>% on cortico-steroids</th>
<th>Disease duration</th>
<th># tender joints</th>
<th># Swollen joints</th>
<th>Disease Activity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westhovens 2009</td>
<td>ABA</td>
<td>51%</td>
<td>7 months</td>
<td>31/68</td>
<td>22/66</td>
<td>6.2</td>
</tr>
<tr>
<td>Breedveld 2006</td>
<td>ADA</td>
<td>35%</td>
<td>9 months</td>
<td>31/68</td>
<td>22/66</td>
<td>6.3</td>
</tr>
<tr>
<td>Emery 2008</td>
<td>ETN</td>
<td>50%</td>
<td>9 months</td>
<td>25/68</td>
<td>17/66</td>
<td>6.5</td>
</tr>
<tr>
<td>Emery 2009</td>
<td>GOL</td>
<td>69%</td>
<td>40 months</td>
<td>28/68</td>
<td>15/66</td>
<td>5.1</td>
</tr>
<tr>
<td>St. Clair 2004</td>
<td>INF</td>
<td>38%</td>
<td>10 months</td>
<td>33/68</td>
<td>21/66</td>
<td>NR</td>
</tr>
<tr>
<td>Tak 2010</td>
<td>RIT</td>
<td>48%</td>
<td>11 months</td>
<td>33/68</td>
<td>21/66</td>
<td>7.1</td>
</tr>
</tbody>
</table>
**Example – Evaluating Consistency in NMA of COPD treatments...**

Are summary estimates from our sources of evidence reasonably similar?

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Direct evidence</th>
<th>Indirect evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>LABA vs Placebo</td>
<td>0.87 (0.79-0.96)</td>
<td>0.91 (0.77-1.06)</td>
</tr>
<tr>
<td>LAMA vs Placebo</td>
<td>0.74 (0.64-0.84)</td>
<td>0.95 (0.80-1.13)</td>
</tr>
<tr>
<td>ICS vs Placebo</td>
<td>0.81 (0.74-0.90)</td>
<td>0.90 (0.81-1.00)</td>
</tr>
<tr>
<td>ICS + LABA vs Placebo</td>
<td>0.72 (0.66-0.79)</td>
<td>0.93 (0.82-1.04)</td>
</tr>
<tr>
<td>LAMA vs LABA</td>
<td>0.91 (0.80-1.06)</td>
<td>0.85 (0.75-1.01)</td>
</tr>
<tr>
<td>ICS vs LABA</td>
<td>0.96 (0.92-1.00)</td>
<td>0.93 (0.81-1.06)</td>
</tr>
<tr>
<td>ICS + LABA vs LABA</td>
<td>0.81 (0.75-0.86)</td>
<td>0.82 (0.72-0.90)</td>
</tr>
<tr>
<td>ICS + LAMA + LABA vs LABA</td>
<td>0.91 (0.75-1.11)</td>
<td>0.75 (0.69-0.96)</td>
</tr>
<tr>
<td>LABA + LAMA vs LAMA</td>
<td>1.07 (0.94-1.22)</td>
<td>1.07 (0.84-1.35)</td>
</tr>
<tr>
<td>ICS + LABA vs LAMA</td>
<td>0.97 (0.93-1.02)</td>
<td>0.97 (0.82-1.14)</td>
</tr>
</tbody>
</table>
Considerations for Reporting Network Meta-Analysis
The PRISMA statement (Moher et al, 2009) provides current guidance to optimize the reporting and transparency of systematic reviews, meta-analyses.

NMA has some distinctions over traditional reviews; a PRISMA extension is currently being developed. Current checklists by ISPOR task force, NICE DSU provide some helpful tips/clues as to what these distinctions are.

What are some of the “additional considerations” to be addressed with this type of review?
Summarizing the extent of included evidence...

- **Traditional reviews**: typically will see # of studies and patients reported in main text of the SR/MA
- **Networks**: most commonly see ‘network’ diagrams; nodes/links sometimes sized to reflect #’s of studies and patients, sometimes just reported with additional text


Summarizing treatment effects from data analysis...

- **Traditional reviews**: most commonly see forest plots with summary estimates and confidence intervals of trial level and summary level findings

- **Networks**: no longer see the study level data given mass of trials, but forest plots still often used to present summary estimates...

Authors may also become creative with tables to present estimates more concisely.

Summarizing additional parameters of potential interest (e.g. Treatment rankings and probabilities...)


Tabular formats are also sometimes used...

Salanti et al provide other possibilities for summarizing this information
Reporting guidance for network meta-analysis is currently in the works in the form of a **PRISMA extension statement**

Will benefit systematic reviewers, technology assessment generators and reviewers, peer reviewers, etc.

**Plans?**
- Conduct of a systematic review of quality of reporting concerns in existing network meta-analyses (COMPLETE)
- Online delphi panel survey (LAUNCHING SHORTLY)
- Face-to-face meeting of ~30 experts in fall 2013 to discuss key issues and develop draft of guidance

**Parties currently involved:**
- Today’s presentors (Ed Mills, Kristian Thorlund, Brian Hutton)
- Other experts: David Moher, Georgia Salanti, Doug Altman, Chris Schmid, Anna Chaimani
Introduction to NMA: Summary
Summary: Overview of Network Meta-Analysis

• “Indirect comparisons” let us compare A and B, *even with no head-to-head data*, through common comparator C and indirect data
  – Expand to more treatments and pooling direct and indirect data using NMA
  – Be sure to explore assumptions (similarity, homogeneity, consistency)

• Indirect comparisons, NMA *commonly used in HTA* where there are multiple treatments of interest and need to establish comparative effectiveness
  – Also used to populate economic models

• There is a need for not only methodologists, but also end-users and decision-makers to develop familiarity with NMA.
How to Use an Article Reporting a Multiple Treatment Comparison Meta-analysis

Edward J. Mills, PhD, MSc
John P. A. Ioannidis, MD, DSc
Kristian Thorlund, PhD, MSc
Holger J. Schünemann, MD, PhD, MSc
Milo A. Puhan, MD, PhD
Gordon H. Guyatt, MD, MSc

CLINICAL SCENARIO
You are seeing a 45-year-old patient for whom, 6 weeks previously, you prescribed paroxetine, a selective serotonin reuptake inhibitor (SSRI), for treatment of generalized anxiety disorder (GAD). The patient reports reduced anxiety, but also insomnia and a reduced interest in sex. You wonder if there is another drug the patient might tolerate or if there will be a more effective treatment for his insomnia and reduced sex drive.

Multiple treatment comparison (MTC) meta-analysis uses both direct (head-to-head) randomized clinical trial (RCT) evidence as well as indirect evidence from RCTs to compare the relative effectiveness of all included interventions. The methodological quality of MTCs may be difficult for clinicians to interpret because the number of interventions evaluated may be large and the methodological approaches may be complex. Clinicians and others evaluating an MTC should be aware of the potential biases that can affect the interpretation of these analyses. Readers should consider whether the primary studies are sufficiently homogeneous to combine; whether the different interventions are sufficiently similar in their populations, study designs, and outcomes; and whether the direct evidence is sufficiently similar to the indirect evidence to consider combining. This article uses the existing Users’ Guides format to address study validity, interpretation of results, and application to a patient scenario.

JAMA. 2012;308(12):1246-1253
www.jama.com
QUESTIONS?
bhutton@ohri.ca